



西南石油大学

# 信息检索与分析

## 第7章 搜索引擎

## 课堂三问

**[过去]**这个工具出现的背景是什么？这个工具是怎么发展的？

**[现在]**如何使用这个工具？对我有何价值？


**[未来]**未来这个工具还会怎样发展？有没有更好的解决方案？

工具仅仅是载体，思想才是课程的灵魂。

努力让每一个走进我课堂的人变成更好的自己



# 本次课程内容

- 学习与搜索
  - 搜索引擎简介（出现的必然性、工作原理）
  - 常用搜索引擎及使用方法，未来发展趋势
- 



# 1.学习与搜索

搜索效率决定了学习效率

# 你每天的困惑，大部分可以通过搜索解决

- 目前大数据领域最权威的专家？研究进展？
- 如何填报高考志愿？
- 公司请代言，老戏骨和小鲜肉，哪个性价比高？
- 广告中炫酷但不知名的字体，如何呈现在你的PPT中？
- 新客户要合作，如何判断对方是不是靠谱？
- 为什么别人几分钟搞定的事情，我却需要大半天？

- 
- 
- **学习首先是一个搜索过程**
  - **要找到合适的对象或资源**

## 2. 搜索引擎简介

### 搜 索 引 擎



诞生的时代背景



发展简史



工作的原理



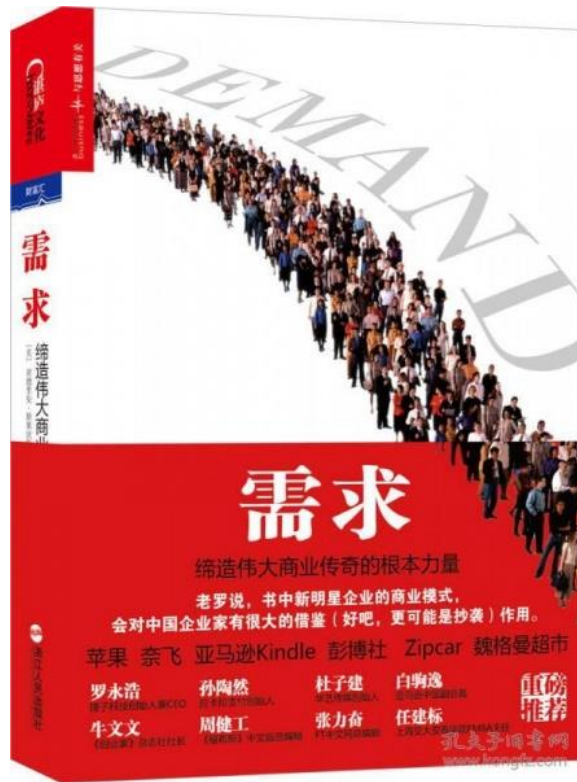
承担的使命

# 一个事物出现的原因是什么？

- 为什么会出现搜索引擎？
- 什么时候，什么背景下出现的？驱动力是什么？
- 如果没有搜索引擎，会出现其它类似的工具么？
- 回到20年前，你能否开发出google 或 百度



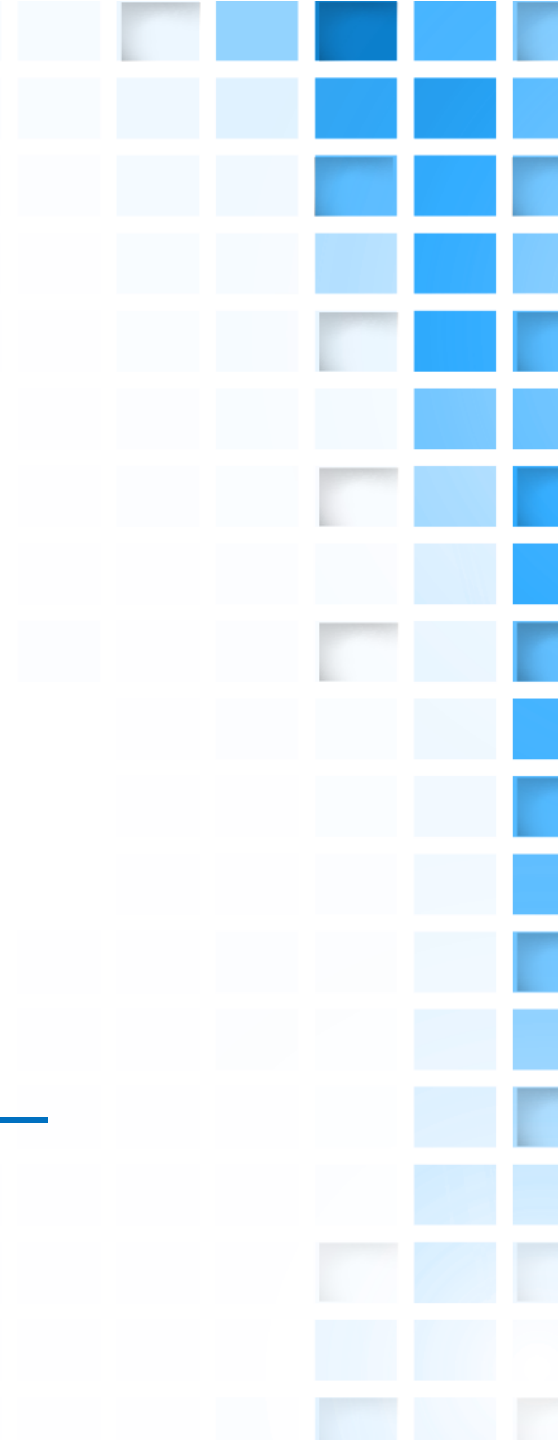
# 什么是需求？



- 需求就是克服种种不便，消灭抱怨
- 每一个抱怨的背后都隐藏着一个未被满足的需求，每一个需求的背后必然隐藏着一个不可忽视的市场



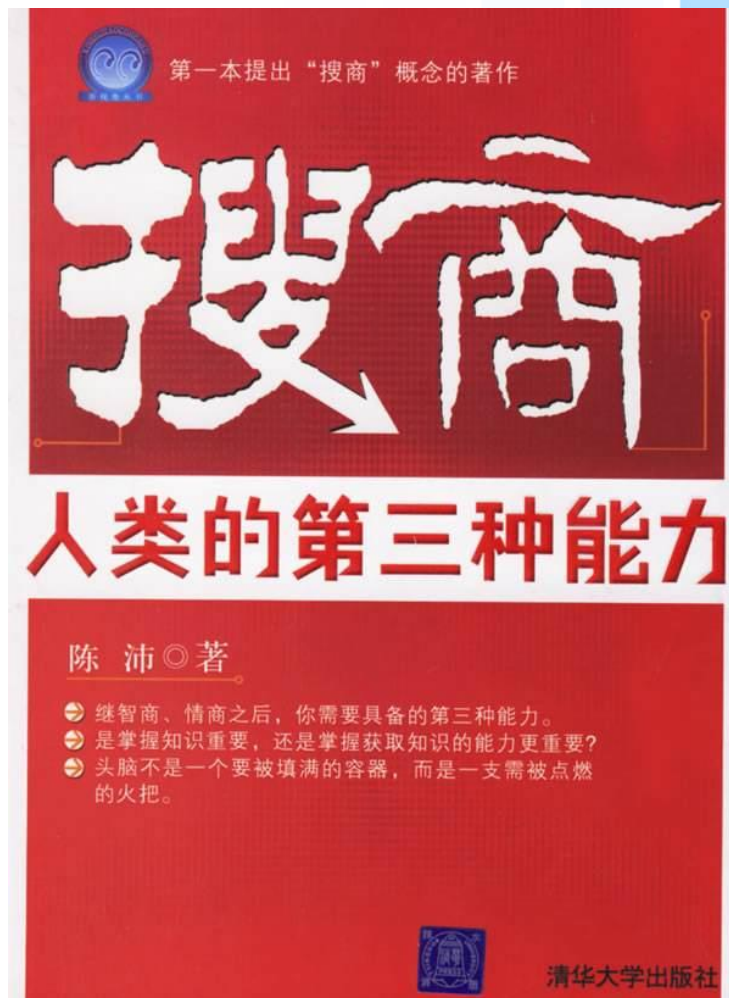
# 人们对搜索引擎的评价

- 世界上最受欢迎的老师
  - 搜索引擎是最好的老师
  - 搜索框正在重塑世界
  - 搜索是最有影响力的学习途径之一
  - 搜索是最大、最好、最具颠覆性的创新之一
  - 改变了信息获取方式和学习方式
- 

# 搜商 智商 情商

同样的问题，越快找到准确的参考资料，你的学习能力就越强，解决问题的能力也越强。

搜商与学习能力和解决问题的能力密切相关。



# 搜索引擎是什么？

- 搜索引擎是指根据一定的策略、运用特定的计算机程序从互联网上采集信息，在对信息进行组织和处理后，为用户提供检索服务的系统。

# 搜索引擎诞生的背景和发展历程

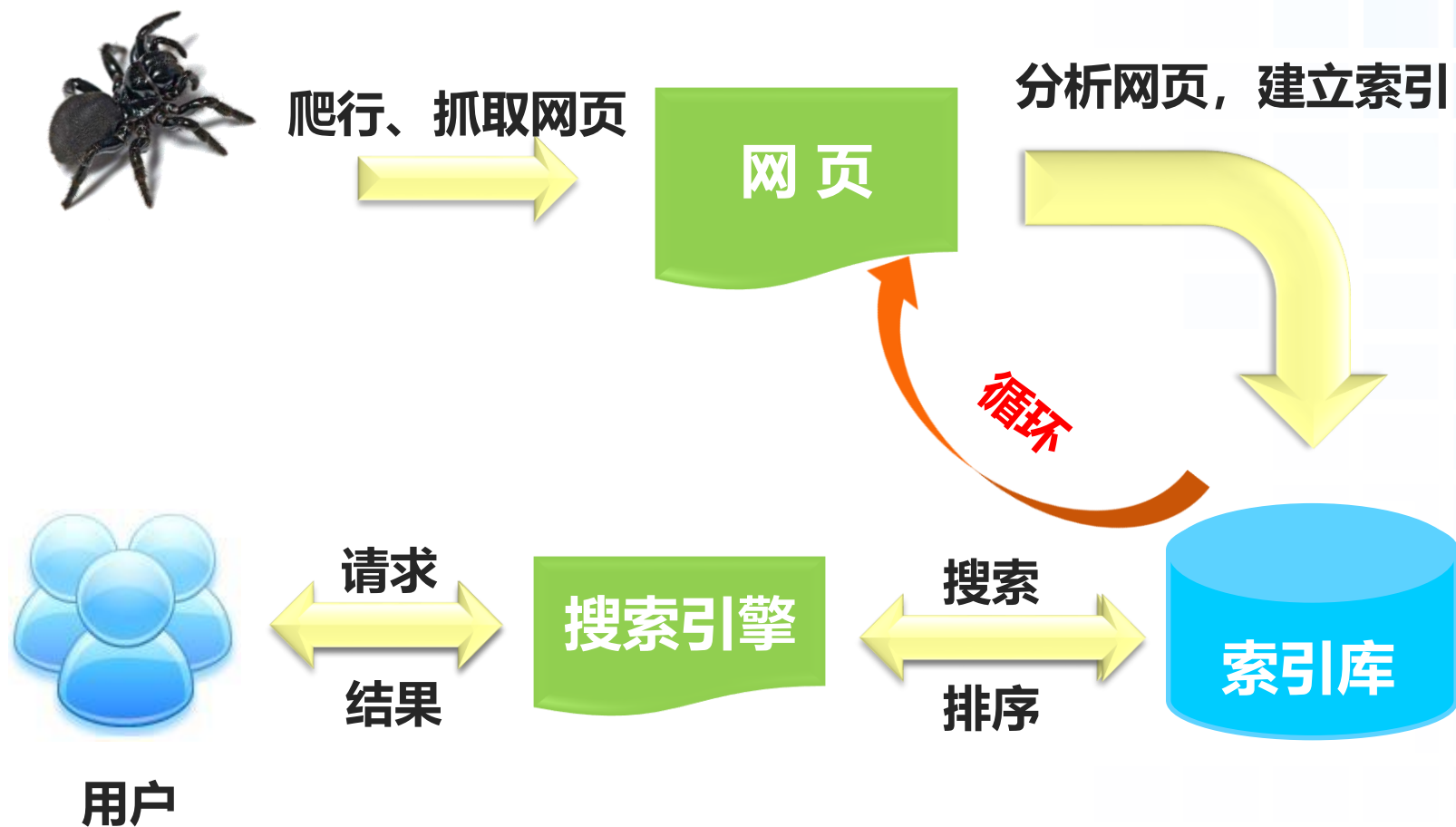
- 互联网的普及是搜索引擎诞生必然因素
- 1994.4 杨致远等创办Yahoo
- 1998.9 Google创立
- 2000.1 李彦宏、徐勇两人创立百度
- 2009年 bing推出
- 2012年8月, 360搜索推出

# 搜索引擎的工作原理



# 搜索引擎的工作原理

教材P273





**提问:不同的搜索引擎之间有何差异?**

**哪些信息可以用来搜索?**

**网页上有哪些信息?**



# 亮点抢先看！杭州亚运会开幕式完成最后一次全要素彩排

## 标题

2023年09月18日 23:04 新华网

A<sup>-</sup> | A<sup>+</sup> | ☆ | 6

新华社杭州9月18日电（记者唐颢宸 徐步云 邱世杰）距离杭州第19届亚运会正式开幕还有5天。9月18日晚，杭州亚运会开幕式最后一次全要素彩排顺利完成。

正文

图  
片

音 频、  
视 频、  
附件...



图为杭州亚运会开幕式彩排画面。（图片由开闭幕式导演组提供）

杭州亚运会开闭幕式总导演沙晓岚介绍，开幕式以“潮起亚细亚”为主题，从中国传统文化中找寻灵感，融入典型的浙江以及杭州符号与意象，借助科技的力量，彰显东方美学的独特吸引力与感染力。整场开幕式的仪式及表演以水为串联，通过“潮起亚细亚”展现新时代中国与亚洲、与世界的文化交融。

# 搜索引擎的分类

- 按照检索范围分：

- 综合搜索引擎

- 资源收录的范围、类型没有限制
    - 百度、谷歌、搜狗

- 垂直搜索引擎

- 资源收录范围限定在某一特定领域或特定类型
    - 小白盘（网盘资源）、鸠摩搜索（电子图书）、DOAJ（OA资源）
    - 图片搜索、音乐搜索、地图搜索、新闻搜索等

- 站内搜索

- 淘宝、抖音、B站、裁判文书网...

# 搜索引擎的分类

- **按照工作机制分：**
  - 独立搜索引擎
    - 独立拥有爬虫、索引库、用户接口等
    - 百度、谷歌
  - 元搜索引擎
    - 通过一个统一的用户界面帮助用户在多个搜索引擎中实现检索，并对检索结果进行优化处理
    - 用户接口、检索代理、结果优化
    - InfoSpace、北斗搜索、360搜索

### 3. 常用搜索引擎及使用方法





# 搜索技巧



# 搜索技巧

## (1) 口语化/问题搜索转化为关键词搜索

**示例：问题搜索 “怎么把我手机上的照片传到电脑里”**

**过多信息，会大幅减少优质结果的数量**

**简化为关键词搜索 “手机照片 传电脑”**

# 关键词如何提取

- 没有明确概念时，可以输入一两个包含范围很大的关键词
- 从结果中寻找“专业词”
- 添加测评，优点，缺点，分类等关键词缩小范围
- 范围缩小后，用范围小的关键词搜索，如

手机——智能手机——安卓手机——华为手机——P30

# 关键词提取技巧

- 参考百度搜索框的提示（热门关键词）
- 不断调整关键词：同义词替换，增加关键词
- 化大为小



# 搜索技巧

## (2) 关键词组合

- 空格、AND、+ (与)

示例：欧洲 新冠肺炎

- OR、| (或)

示例：金融危机 OR 金融风暴

- - (非) 去除不需要的内容

示例：信息检索 - 文库

减号前面必须是空格，减号后面没有空格



# 搜索技巧

## (3) 搜索短语和句子

**示例：***"knowledge management"*

# 搜索技巧

## (4) 限定在特定网站中 (site)

**示例：复试分数线 `site:swpu.edu.cn`**

### **重名查询**

- 作用
  - 替代站内搜索
  - 指定网站搜索，缩小范围，提供精准结果
  - 搜索不能直接访问的网站；死网站；绕过注册

## 搜索技巧

### (5) 把检索结果限制在网址里面 (inurl)

https://cet6.koolearn.com/zhenti/10.html

域名

网页名称

网页在Web服务器中的存储路径

**示例:** *inurl:cet4*

**真题** *inurl:cet6*

*vlog inurl:jiaocheng*

*excel inurl:jiqiao*

*inurl:download inurl:ebook intitle:数学之美*

# 搜索技巧

## (6) 限定文件类型 (filetype)

- 适用于专业文档的搜索，如pdf、ppt、doc、jpg等

用法：关键词+空格+filetype:文件格式

- 示例：

- *2022考研英语 filetype:pdf*

- *简历 filetype:doc*



## 搜索技巧

### (7) 限定在标题中 (intitle/title)

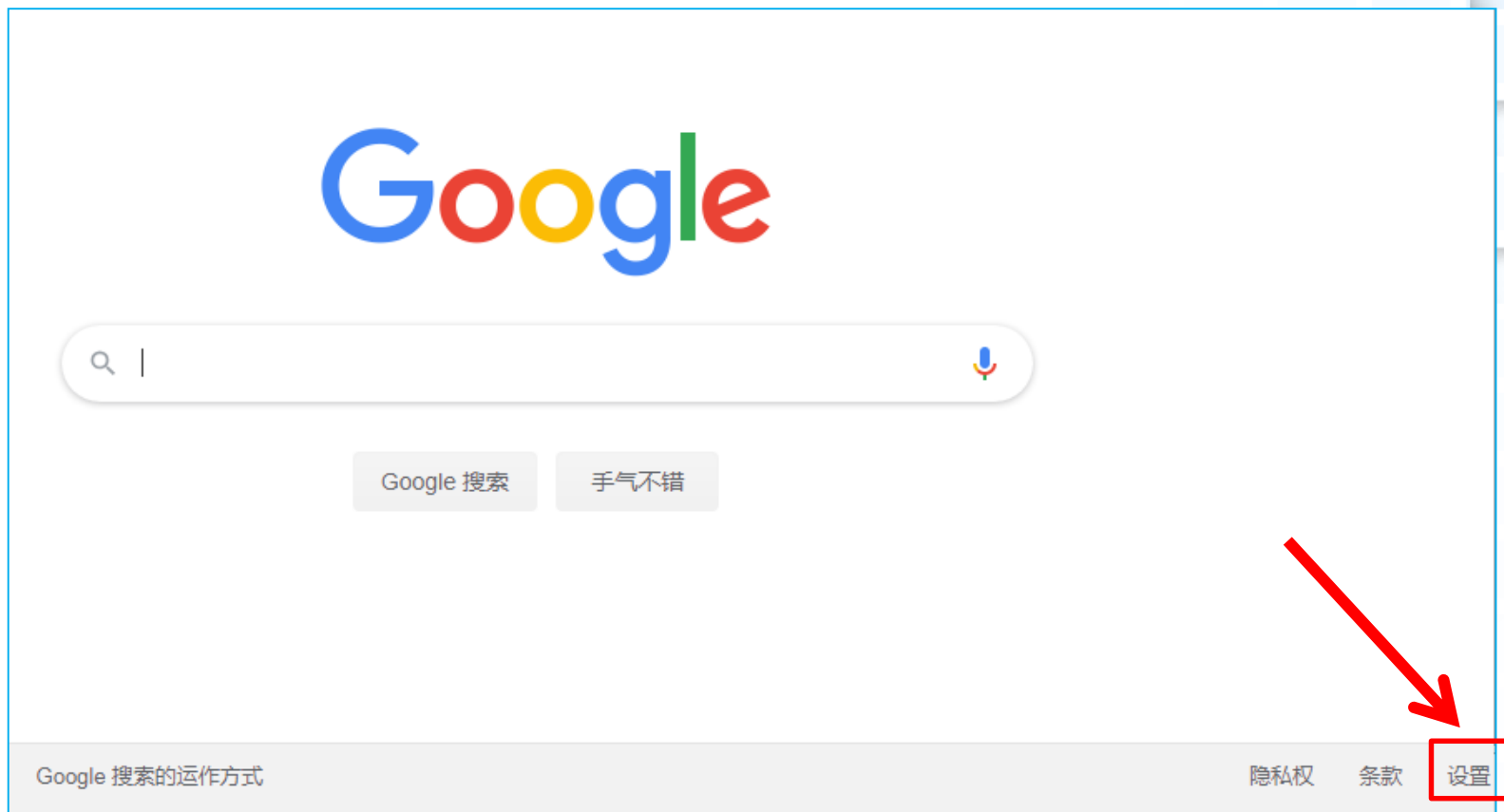
示例：

**intitle:人工智能**



# 高级搜索

# 高级搜索的进入方式-google





# Google 高级搜索界面

使用以下条件来搜索网页...

以下所有字词：

与以下字词完全匹配：

以下任意字词：

不含以下任意字词：

数字范围：从

到

在搜索框中执行以下操作。

输入重要字词： 杨山鸭梨

用引号将需要完全匹配的字词引起：“鸭梨”

在所需字词之间添加 OR： 批发 OR 特价

在不需要的字词前添加一个减号： -山大、-“刺梨”

在数字之间加上两个句号并添加度量单位： 10..35 斤、300..500 元、2010..2011 年

然后按以下标准缩小搜索结果范围...

语言：

任何语言

查找使用您所选语言的网页。

地区：

任何国家/地区

查找在特定地区发布的网页。

最后更新时间：

任何时间

查找在指定时间内更新的网页。

网站或域名：

搜索某个网站（例如 wikipedia.org），或将搜索结果限制为特定的域名类型（例如 .edu、.org 或 .gov）

字词出现位置：

网页上任何位置

在整个网页、网页标题、网址或指向您所查找网页的链接中搜索字词。

安全搜索：

安全搜索滑块：从 关闭 到 适中 到 严格

指定安全搜索针对色情内容的过滤等级。

文件类型：

任意格式

查找采用您指定格式的网页。

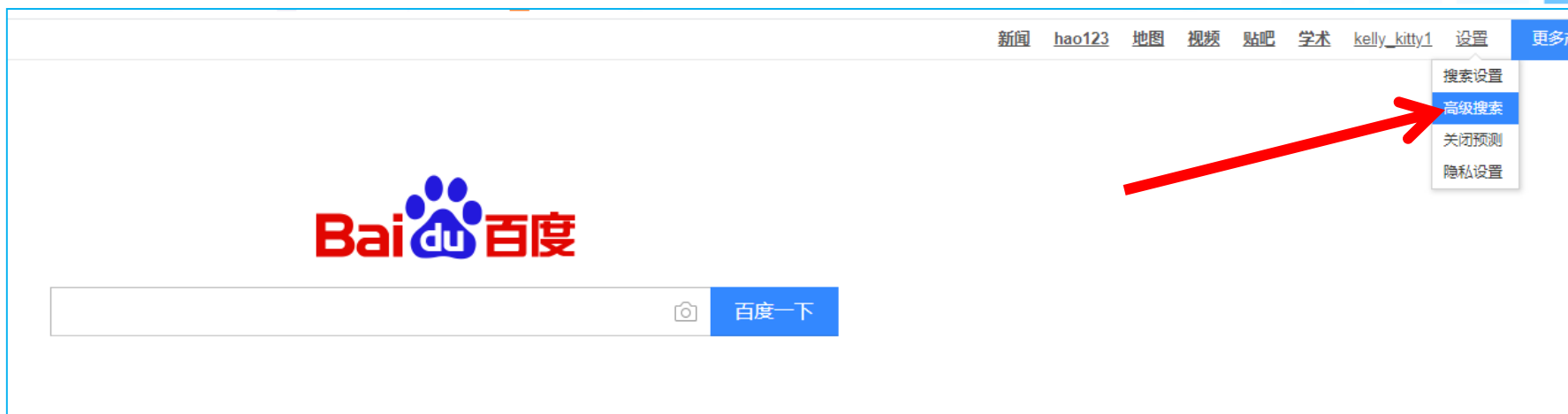
使用权限：

不按许可过滤

查找可自己随意使用的网页。

高级搜索

# 高级搜索的进入方式-baidu



# 百度高级搜索

**搜索结果：**包含以下**全部**的关键词

包含以下的完整关键词：

包含以下任意一个关键词

不包括以下关键词

**时间：**限定要搜索的网页的时间是

全部时间 ▼

**文档格式：**搜索网页格式是

所有网页和文件 ▼

**关键词位置：**查询关键词位于

☒ 网页的任何地方 ☐ 仅网页的标题中 ☐ 仅在网页的URL中

**站内搜索：**限定要搜索指定的网站是

例如：

baidu.com

高级搜索

# 高级搜索的进入方式-360



# 360高级搜索

搜索设置

高级搜索

搜索时间：限定要搜索的网页时间

全部时间 ▼

文档格式：搜索网页格式

所有网页和文件 ▼

站内搜索：限定要搜索定制的网站

例如：so.com，目前不支持多网站同时站内搜索

高级搜索

# 部分应用举例

- 查找新冠肺炎诊疗方案
- 利用site快速找出某个网站上特定类型的文件
- 直接搜索 Site:swpu.edu.cn 会得到什么结果?

# 利用搜索引擎解决问题的思路

- 搜索意识，逆向思维与换位思考
- 哪里会有这些信息？谁会知道？\*\*\*\*\*
- 怎么才能搜索到，去哪里搜会最高效？（网站、社交平台、专家）
- 是否会在网上出现？哪里会收录？哪个搜索引擎会收录？
- 发布者会怎么表述

# 搜索的未来

## 搜索：开启智能时代的新引擎



- 交互搜索，无屏幕搜索（智能音箱）
- 基于位置、场景（室内搜索）
- 理解用户需求（根据用户历史信息），  
搜索酒店并直接预订
- 搜索看成是链接虚拟世界和现实世界的“铰链”，通过崭新的交互方式（语音问答）、智能内核（人工智能）以及对现实更强理解力（物联网，大数据），搜索将成为你助理，拓展你的能力（注意力、记忆、智能等）。搜索将从提供信息、到智慧筛选再到提供建议和促发行动。





**THANK YOU**